T Level Technical Qualification in Digital Business Services

Occupational specialism assessment (OSA)

Data Technician

Task 2

Assignment brief

v1.2: Additional sample material
24 November 2023
603/6902/4

T Level Technical Qualification in Digital Business Services Occupational specialism assessment (OSA) (603/6902/4)

Data Technician

Assignment brief

Task 2

Contents

About this assignment

Introduction

This occupational specialism assessment (OSA) is set by NCFE and administered by your provider during a 3 week window. It contains 4 separate tasks which will be completed one after the other during this assessment window.

All 4 tasks will be completed under supervised conditions.

You must complete all tasks in this assignment independently. You are required to sign a declaration of authenticity to confirm that the work is your own. This is to ensure authenticity and to prevent potential malpractice and maladministration. If any evidence was found not to be your own work, it could impact your overall grade.

You will be given a copy of the assignment brief and any relevant supporting information with each task, so you do not have to memorise any information.

Timings

You have a total maximum time of 29 hours to complete all tasks within this assignment, and each task has the following number of hours to complete it:

Task 1 – 5 hours

Task 2 – 10 hours

Task 3 – 8 hours

Task 4 – 6 hours

Individual tasks must be completed within the timescales stated, but it is up to you to decide how long you spend on each part of the task, therefore you should manage your time appropriately.

Details on the separate marks available are provided in each task.

You should attempt to complete all of the tasks.

Read the instructions carefully.

Performance outcomes (POs)

Marks will be awarded against the skills and knowledge performance outcomes (POs) as follows:

Task 2

This task is divided into 2 parts (part A and B) and carries a total of 52 marks.

These are divided between the following POs:

- PO1: Source, organise and format data securely in a relevant way for analysis (16 marks)
- PO2: Blend data from multiple sources (20 marks)
- PO3: Analyse structured and unstructured data to support business outcomes (16 marks)

Scenario

Data analytics is surging in popularity; a core reason for this is that data can be used to make strategic and impactful business decisions. The data collected allows businesses to identify gaps in the market which in turn can create further business opportunities and growth.

Market trends regularly change, and it is important this is tracked through data analytics. Start-up Analytics is a company that specialises in data trends and market change. The company has become popular over recent years as a result of more data being tracked and stored through digital media. External clients use Start-up Analytics to create business strategies based on data analysis. With the increased use of internet-connected devices, there are more opportunities to work with large data sources to predict future forecasts.

About you and your employer

You are the junior data technician for the research and development department and work for a company called Start-up Analytics who specialise in providing data relevant to new business ventures looking to start operations within England. You work in a small team of 5 and are led by the data analytics manager, Charlie Johnson. Your role includes analysing existing data provided by the client and collecting, cleaning and analysing data from a range of external datasets available from websites, social media and demographic sources. As a junior data technician for a business analytics agency, you report to the experienced data analytics manager to produce reports which are used for increasing sales and supporting start-up businesses.

Charlie Johnson works closely with Muhammad Mia who is the business consultant for Start-up Analytics. Together they advise clients on how to focus on key areas for success. Muhammad will regularly hold meetings with Charlie and request reports on trends and growth around the country, these reports and basic information keep Muhammad informed to allow for the client to receive a clear business strategy. As the data analytics manager, Charlie will receive reports directly from you and you must work with raw data to provide conclusions into new business opportunities.

About the client

Your client works in the education sector and is aiming to be one of the largest providers in the country offering private qualification training, they are considering branching out into apprenticeship qualifications. The client has limited knowledge of the apprenticeship sector and is unsure where to open their hubs and which apprenticeship qualifications to focus on.

The clients new vision statement is: “To provide ‘earn while you learn’ opportunities through apprenticeships”.

Their objectives for the coming year are:

- open multiple training hubs across the country which provide 12 to 18 month apprenticeship qualifications
- highlight the most popular industries, locations and sectors to start apprenticeships within, and use their already large training resources to work with current customers
- highlight and work within well populated and high-salaried locations to minimise lead time and close sales
- build their credibility throughout the country to be the number one training provider
- promote equality within the education sector with a focus on reducing the gender pay gap

The brief

The client has selected Start-up Analytics to help determine suitable locations for training hubs across England. They have a range of resources at their disposal, but are unsure on what location, industry and sector to focus on.

As a junior data technician, you will provide the client with information to help them succeed in meeting their business objectives. This will be done by sourcing an array of appropriate information available online, including data from the Office of National Statistics (ONS). You will source up to date data on the types of apprenticeships available, the most sought-after qualifications and industries to start business within. The client requests the proposal before starting the procurement of buildings and staff. The client is also interested in data insights to support working with young and vulnerable people in the education sector.

The average age of the clients' customers is 16 to 30 years of age, they understand how GDPR can impact how they use their data to make profit within the education sector. The client believes that the digital sector should be their main focus. They believe the sector has the highest salaries and that salaries are higher in the south of England. However, they have no data to support this.

The client has told you the following:

- on average, customers are under the age of 30
- the current qualifications the business focus on are accountancy, customer service and administration
- they have one office, in Manchester, England
- they currently offer in-house and location-based training, but are interested in moving towards a remote and virtual model of training

The client has provided you with data focusing on salaries over the years, gender pay gap, pay by industry and salary by geographical location. Use this dataset, business objectives and what the client has told you about their business to justify your decisions throughout the project.

Your role

Throughout this project you are required to collate appropriate data from a variety of sources, both internal and external. You will need to judge how useful the data is towards the company objectives; you will also need to bring together multiple datasets into a combined location to provide correlation and solution. The client has not cleansed any data and it includes a range of structured and unstructured data; it is up to you to provide clarity for the client based around their objectives.

You are required to identify trends or patterns you see in the data you collect; once data has been cleansed, transformed and modelled, you are required to provide a data dashboard to summarise the data for the client.

To support future analysis, you have been asked to log the types of data formatting used and the methods for verification and validation of your data. It is critical that security measures are considered and that you are in line with relevant legislation when using all available resources, this shall provide reason behind your insights and recommendations. This will also be important for the client to understand the rationale and whether the return on investment will be made when taking on apprenticeships. The proposal must be data driven and allow for the client to explore possible options for growth.

Task 2

Time limit and marks available

Maximum time allowed: 10 hours (you can use this time how you want during each session, but task 2 must be completed within this time limit).

(52 marks)

Instructions for students

The client intends to open training hubs within major locations across the country. The client wants to allow for in-house training for its customers but wants to understand how many potential customers they could attract before they purchase a building. They are currently undecided which locations and sectors to open their respective hubs in and want to use a combination of their in-house data and publicly available data to inform their decision.

Charlie Johnson has provided you with some internal customer and sales data, along with external datasets.

Part A

The client wishes to open 10 training hubs across England. Each hub has potential to offer a different apprenticeship qualification which will support sector demand in that region.

Charlie has asked you to join the external data into one single clean dataset. Make sure the single dataset has appropriate variables which reflect the client's business objectives, as it will eventually help to create a dashboard for the client. Your client does not require any information for those above the age of 30.

Once cleansed and validated, you must suggest 10 locations the client should choose and which apprenticeship qualification each location should deliver based on the number of jobs available for a certain occupation. The recommendations for apprenticeship qualifications should be appropriate to the needs of each specific location. You should explain why you chose a certain apprenticeship qualification and exclude any industries where the gender pay gap is above the median compared to other occupations/industries. The final dataset should also include any calculations which may help you design a dashboard.

Charlie would like you to keep a log of your progress and any decisions you make.

This log must include:

- which variables you consider relevant to the business objectives and why
- errors you have found in the datasets
- ways you have validated the data
- which columns you feel are appropriate to the business objectives and why
- the primary keys for each dataset
- data you have removed and why
- the occupations you considered significant and why
- any calculations and aggregations you have applied to the data
- how you reformatted the data to be joined to the client's internal data

Include any code or formulas you used to automate the above tasks.

Part B

For this part of the task, the internal data received from the client has been exported from their relational MySQL database. They plan to upload the single dataset you created in part A to their infrastructure. The dataset will include the following tables:

- annual_gross_pay_by_occupation.xls
- apprenticeship_demographics.xls
- current_learners.xls
- median_hourly_earning_by_age.xls
- your available apprenticeship qualifications dataset

In addition to part A**,** Charlie has asked you to write a separate additional section in your log. You must:

- describe the normalization form of this new dataset, giving a clear explanation of your reasons
- identify the primary, alternate and foreign keys in each table – write a sentence for each key describing why you have identified it as such
- explain how you reformatted the data to be joined to the external data
- explain how you manipulated date of birth to a format appropriate to the context
- provide a data validation template for each column in your new table which includes **data types and constraints**
- explain how you removed any variables from the internal datasets that is **not** applicable for your analysis
- include any code or formulas you used to automate the above tasks

Resources

You will have access to the following resources, plus the original brief:

- task 2 datasets:
  - current_learners.xls
  - median_hourly_earning_by_age.xls
  - annual_gross_pay_by_occupation.xls
  - apprenticeship_demographics.xls
  - average hourly pay (excluding overtime) in the regions and devolved countries of the UK.xls
  - average weekly earnings by industry.xls
  - prov– Age by Occupation SOC10 Table 20.12 Gender pay gap 2021.xls
  - prov – Home Geography Table 8.12 Gender pay gap 2021.xls
  - prov-– Home Travel to Work Area Table 12.12 Gender pay gap 2021.xls
  - prov – Work Region PubPriv Table 25.7a Annual pay – gross 2021.xls
  - prov – Work Region PubPriv Table 25.9a Paid hours worked – total 2021.xls
  - prov – Work Region PubPriv Table 25.12 Gender pay gap 2021.xls
- software applications to clean and blend data (for example, Microsoft or Google)
- word processing software (for example, Microsoft or Google)

Note: You will not have access to the internet during this task

Evidence required for submission

- single joined dataset
- decision log of processes and steps you took as described in the instructions for both part A and part B

Document information

The T Level Technical Qualification is a qualification approved and managed by the Institute for Apprenticeships and Technical Education. NCFE is currently authorised by the Institute to develop and deliver the T Level Technical Qualification in Digital Business Services.

Owner: Head of Assessment Design

Change History Record

Version	Description of change	Approval	Date of Issue
v1.0	Additional sample material		01 February 2023
v1.1	Minor amends to task 2 to improve clarity		01 September 2023
v1.2	Sample added as a watermark	November 2023	24 November 2023